Özgür, C. (2024). "Relevance Vector Machines for Index Direction Predictions: An Application on Borsa Istanbul", Eskişehir Osmangazi Üniversitesi İİBF Dergisi, 19(2), 594 – 610.
Doi: 10.17153/oguiibf.1400125

Başvuru: 04.12.2023 Kabul: 04.02.2024 *Araştırma Makalesi/Research Article*

Relevance Vector Machines for Index Direction Predictions: An Application on Borsa Istanbul

Cemile Özgür[1]

Endeks Yönü Tahmininde İlgililik Vektör Makineleri: Borsa İstanbul Üzerine Bir Uygulama

Öz

Bu çalışma, İlgililik Vektör Makineleri (İVM) ile sıklıkla uygulanan Destek Vektör Makineleri (DVM) ve Ridge Regresyonunun endeks tahmin performansını araştırmaktadır. Algoritmaların bir gün sonrası tahminlerinin elde edilmesi amacıyla Borsa İstanbul'un BIST Banka ve BIST Mali endekslerinin günlük fiyat serileri kullanılmıştır. Hesaplanan performans ölçütlerine göre İVM, BIST Banka'nın her iki periyodunda da çoğunlukla en iyi ölçütleri sağlamıştır. BIST Mali endeksinin en iyi performans ölçütlerini DVM elde etmişken, İVM'nin ölçütleri en iyiden çok uzakta değildir. Genel olarak sonuçlar, İVM'nin endeks yönü tahmininde uygulanabilirliğini ve DVM'nin iyi bir rakibi olma potansiyeline sahip olduğunu belirtmektedir.

Anahtar Kelimeler: Endeks Tahmini, Destek Vektör Makineleri, Ridge Regresyonu, İlgililik Vektör Makineleri

JEL Kodları: C22, C45, C53

Relevance Vector Machines for Index Direction Predictions: An Application on Borsa Istanbul

Abstract

This study investigates index prediction performance of Relevance Vector Machines (RVM) and frequently applied Ridge Regression and Support Vector Machines (SVM). Daily prices of BIST Banks and BIST Financials indices of Borsa Istanbul are used to obtain one-day-ahead predictions of the algorithms. According to estimated performance measures, RVM yielded mostly the best metrics in both periods of BIST Banks. While SVM obtained the best performance metrics on BIST Financials index, metrics of RVM were not far from the best. Overall, the results indicate the applicability of RVM in predicting index directions and has a potential to be a good rival of SVM.

Keywords: Index Prediction, Support Vector Machines, Ridge Regression, Relevance Vector Machines

JEL Codes: C22, C45, C53

Araştırma ve Yayın Etiği Beyanı	Bu çalışma bilimsel araştırma ve yayın etiği kurallarına uygun olarak hazırlanmıştır.
Yazarların Makaleye Olan Katkıları	Yazar 1'in makaleye katkısı %100'dür.
Çıkar Beyanı	Yazarlar açısından ya da üçüncü taraflar açısından çalışmadan kaynaklı çıkar çatışması bulunmamaktadır.

[1] Dr., Bağımsız Araştırmacı, ozgurcemile@yahoo.com

1. Introduction

Stock market prediction is a popular as well as a highly challenging research topic in economics and finance. Accurate stock market prediction is vital for the market participants in order to develop hedging or profitable investment strategies. In predicting prices, price direction or returns of stocks/stock markets, various econometric methods such as ARIMA and GARCH are commonly employed. However, there are several critics directed to these traditional methods stemming from the complex, nonlinear and dynamic nature of financial time series. Especially in the last 20 years, an increasing number of machine learning (ML) algorithms e.g., the Neural Networks and Support Vector Machines (Vapnik, 1998), are frequently applied to be able to improve accuracy of the prediction tasks.

While there is a growing literature on stock/stock market prediction using various ML algorithms, there is still a few papers focusing on the prediction of different indices of Borsa Istanbul. Moreover, it is even harder to find a research paper employing several econometric and/or ML methods to predict an index of Borsa Istanbul other than BIST100 or BIST30 indices. As a result, this research investigates price direction prediction performance of two underexplored stock indices (BIST Banks and BIST Financials) of an Emerging Stock market, Borsa Istanbul, by employing various ML algorithms as well as the traditional ARMA-GARCH.

The second main contribution of this paper is that, while SVM is a frequently applied algorithm, the number of research papers focusing on the prediction of stock market price or price direction that employ Relevance Vector Machines (RVM) (Tipping, 2001) is quite few. Moreover, it is even harder to find one (if not possible) employing RVM to predict one of the indices of Borsa Istanbul, other than BIST100 or BIST30.

Furthermore, one of the gaps in the literature is that there is still a room in investigating prediction performance of linear methods such as the Ridge (Hoerl & Kennard, 1970) regression. Ridge regression incorporate a term (ℓ_2 norm) into the classical linear regression equation in order to penalize highly correlated or redundant variables by shrinking their coefficients. By performing shrinkage or regularization, variance of a model can be reduced with an increase in bias as a trade of. Like RVM, it is also difficult to find research papers that focus on the indices of Borsa Istanbul and compare the prediction performance between the linear and nonlinear ML algorithms.

Finally, one of the other contributions of this paper is that the existing literature employing ML algorithms mainly use the default parameters or k-fold cross-validation to train and select the optimal hyper-parameters of the algorithms which ignores time series properties of the stock or stock index prices. In contrary, in this research time-series cross-validation is applied to train the ML algorithms and select the optimal hyper-parameters. Additionally, only lagged values of the indices are used as the input variables to the ML algorithms applied in this research. As stated by Avcı (2007) and Oztekin *et al.* (2016), the selection of input variables is critical for ML algorithms, since an increasing number of model inputs may reversely effect prediction performance of an algorithm by decreasing its simplicity and parsimony.

The main sections of this paper are structured as follows: Section 2 reviews the literature especially focusing on the indices of Borsa Istanbul, Section 3 briefly defines the employed prediction methods as well as the performance evaluation metrics. Section 4 introduces the data, defines some of the methodological steps and reports the empirical findings. Section 5 discusses the findings and concludes the paper.

2. Literature Review

There is a vast literature on stock market prediction that employ various econometric methods. Nowadays, an increasing number of machine learning algorithms are applied to forecast price, price direction and/or returns of stocks, stock indices and/or the other financial assets. For example; Huang and Wu (2008) employed RVM to forecast one-day ahead returns of four indices (NASDAQ, NK225, TWSI and KOSPI) after applying a wavelet decomposition to the data. Prediction performance of the wavelet decomposed RVM is compared with the traditional GARCH as well as the wavelet-kernel SVM and SVM. According to the obtained performance metrics, authors argue that the proposed wavelet decomposed RVM yields promising results in terms of reducing the prediction errors.

Ballings *et al.* (2015) predicted one-year ahead price direction of 5767 stocks by using three ensemble algorithms (AdaBoost (AB), Kernel Factory (KF), Random Forest (RF)) additional to Neural Networks, SVM, Logistic Regression and KNN. Moreover, the researchers employed a total of 81 model input variables consisting company specific information as well as various macroeconomic and financial statement variables. According to the research results, RF, SVM and KF are ranked as the top three performers, respectively. As a result, the researchers expressed the importance of employing ensemble methods, especially RF, in predicting stock price direction.

On the other hand, Nikou *et al.* (2019) applied various machine and deep learning algorithms to predict daily closing prices of an exchange-traded fund named iShares MSCI United Kingdom. The researchers compared prediction performance of Artificial Neural Network (ANN), SVR, RF and LSTM with the commonly employed MAE, MSE and RMSE metrics. The research results indicate that LSTM and SVM are the best and the second-best performers, respectively. Since this paper is focused on the indices of Borsa Istanbul, the following paragraphs briefly summarizes the works aiming to forecast at least one of the indices and/or stocks of Borsa Istanbul by also employing various ML algorithms. Moreover, one can consult to the works of Henrique *et al.* (2019), Kumbure *et al.* (2022) and Sahu *et al.* (2023) for a worldwide review of literature.

One of the early research papers that employ machine learning to forecast one day ahead price direction of BIST100 index is the Diler (2003)'s work in which he applied the ANN backpropagation algorithm. The author used one-day lagged values of various technical indicators as well as the returns of the index as model inputs. The research results indicate 60.81% success in test and 59.67% in all data direction forecasts.

Avcı (2007) also employed the ANN algorithm to forecast daily as well as seasonal returns of BIST100 index for the period of January 1996 and June 2005. The researcher employed fourteen model input variables consisting lagged index returns, change in volume, moving averages for the index returns and for the volume. According to the estimated performance metrics and applied sensitivity analysis, the researcher argues that the performance of ANN is not consistent among different test sample periods. Moreover, the forecasting frequency (such as daily, monthly, or seasonal) and the selection of relevant input variables are the other two factors that one needs to consider when employing the algorithm.

Moreover, Boyacıoğlu and Avcı (2010) applied Adaptive Network-Based Fuzzy Inference System (ANFIS) algorithm to predict monthly returns of BIST100 index for the period of January 1990 – December 2008. The researchers employed closing prices of three international indices (DAX, DJI, BOVESPA) and various macroeconomic variables as input variables to the algorithm. According to the estimated performance metrics, it is argued that the applied algorithm (ANFIS) is an appropriate method to forecast stock market returns.

In another work, Kara *et al.* (2011) employed ANN and SVM to predict daily price direction of BIST100 index by using ten technical indicators as model inputs for the period of 1997 - 2007. According to the estimated model specific average prediction performance percentages, the authors indicate that the ANN algorithm (75.74%) is significantly better than SVM (71.52%) in predicting index direction.

Akcan and Kartal (2011) predicted prices of seven stocks listed in ISE Insurance index by using the ANN algorithm for varying time horizons (fifteen days, one month, one and half months and two months). The researchers employed four macro and eight microeconomic variables as input variables to ANN. The results are evaluated with the MAE and MAPE metrics and compared across the four forecast horizons. Overall, the predictions of the algorithm are found quite successful, especially for the forecast horizons of up to one month.

Furthermore, Oztekin *et al.* (2016) employed three algorithms (ANN, ANFIS and SVM) to predict the daily percentage change in the BIST100 index for the period of 2007 - 2014. The researchers used 10-fold cross-validation to train the models with several input variables, such as the percentage changes in gold prices, Nasdaq Composite and FBIST Bond indices. In terms of the estimated accuracy metrics, the researchers argue that their forecasting approach outperformed the previous works on BIST100 index (e.g., the works of Diler (2003) and Yümlü *et al.* (2005)).

Gündüz *et al*. (2017) predicted daily price direction of three stocks (GARAN, THYAO and ISCTR) of Borsa Istanbul by employing the gradient boosting machine (GBM) and logistic regression with and without feature selection. For this purpose, the authors employed 5860 features consisting stock specific technical indicators as well as indicators of other BIST100 stocks and the market. Performance of the applied algorithms are compared with the F-measure and accuracy metrics. According to the obtained metrics, applying feature selection improved the classification performance of the algorithms. On the other hand, including the other BIST100 stock indicators as features improved the performance of the classifiers for all the three stocks.

Using ten technical indicators as input variables, Pabuçcu (2019) applied ANN, SVM and Naïve Bayes algorithms for a purpose to predict daily price direction (positive or negative) of BIST100 index during the period of 2009 – 2018. The author compared the classification performance of the algorithms with the precision, sensitivity, and F-score measures. Considering the estimated performance measures, the author indicates that while all the applied algorithms can be employed to predict the index direction, the ANN algorithm outperformed the rest with a better classification performance.

One of the other works in index prediction is the Kartal (2020)'s paper in which he applied the SVM algorithm to predict monthly directions of four stock indices (BIST100, S&P 500, DAX, and NIKKEI 225). The researcher employed various macroeconomic and index data as input variables to SVM, such as; EUR/USD exchange rate, price of Brent oil and Gold as well as the indices of MSCI World, MSCI ACWI and MSCI Europe. Prediction performance of SVM (using the linear kernel) is evaluated with the metrics of mean absolute error (MAE), accuracy, F-score, ROC and the Kappa statistic. The research results indicate a direction classification accuracy of %62,86 for BIST100, %81,523 for S&P500, %81,579 for DAX and %79,483 for NIKKEI 225.

On the other hand, Filiz *et al.* (2021) applied several machine learning algorithms (e.g. SVM, KNN, logistic regression and ANN) to classify the change direction of BIST100 index for the period of 01.01.2006 – 01.12.2020. The researchers employed a total of twenty-eight features consisting twenty-five international stock indices, USD/TL, and EUR/TL exchange rates as well as gold price as inputs to the algorithms. Additionally, a feature selection is performed by using the CfsSubset algorithm. Performance of the applied algorithms are compared with the accuracy, RMSE, ROC, Matthew correlation coefficient (Mcc) and Kappa statistics. According to the estimated metrics, logistic regression obtained the best classification accuracy of 70.6% before the feature selection is applied. After the feature selection step, SVM PUK core algorithm outperformed the rest by yielding a classification accuracy of 71.9%.

Aksoy (2021) employed ANN, Classification and Regression Tree (CART), and K-Nearest Neighbour (KNN) algorithms to forecast three months ahead price direction of five stocks (ARCLK, EREGL, SISE, TOASO and TUPRS) listed in Borsa Istanbul by employing their financial statements as well as various macroeconomic variables. The results of the research show 98.05% classification accuracy for ANN, 96.10% and 92.20% accuracy percentages for CART and KNN algorithms, respectively. Ünlü *et al.* (2021) also aimed to predict the daily direction of BIST100 index. For this purpose, the researcher employed several technical indicators as input variables to the SVM that use various kernel functions. Moreover, Random Forest algorithm is also employed as a feature selection tool. Overall, the estimated train and test set metrics indicate a prediction performance in favour of SVM especially after the feature elimination.

Özgür and Sarıkovanlık (2022) applied single (Random Forest, XGBoost and ANN) and hybrid machine learning algorithms to forecast returns of BIST100 and NASDAQ indices. Additional to developing a unique hybrid forecasting framework, the authors also developed a trading strategy to test the economic impact of return forecasts of the employed single and hybrid algorithms. According to the estimated performance metrics, the authors argue that the proposed novel hybrid algorithms are able to yield quite promising out of sample test results compared to the rest of the algorithms applied in the research.

In one of the recent works, Kılıç *et al*. (2023) predicted the sign of twenty-six stocks listed in BIST30 index of Borsa Istanbul by using the 5-min intraday stock data for the year of 2018. The researchers employed various machine learning algorithms (logistic regression, KNN, RF, SVM using different kernels, decision tree and naïve Bayes) including PCA for dimension reduction. Results of the research shows a statistically significant sign prediction performance of ML algorithms in nine out of twenty-six stocks compared to the employed random predictor as the benchmark.

3. Methodology

This part briefly defines the algorithms applied to predict price direction of the two stock indices as well as the performance metrics employed to evaluate and compare the accuracy of the forecasts.

3.1. ARMA-GARCH

In finance, it is a well-known fact that financial time series have the properties of being correlated with their own lagged observations (auto-correlation) as well as with their absolute or squared values (heteroskedasticity). As a result, a combination of Autoregressive Moving Average (ARMA) and the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) (Engle, 1982; Bollerslev, 1986) are the frequently applied models in time series forecasting tasks. ARMA(p,q) is defined as follows:

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{k=1}^{q} \theta_k \varepsilon_{t-k} + \varepsilon_t \quad (1)$$

where y_t defines the dependent variable observed at time *t*. y_{t-i} defines the lagged values of y_t for the lags of $i = 1,2,3,\dots,p$. ε_t and ε_{t-k} are the residual and the lagged residual terms ($k = 1,2,3,\dots,q$), respectively. On the other hand, the conditional variance equation of GARCH*(P,Q)* can be defined as follows:

$$\sigma_t^2 = \omega + \sum_{m}^{P} \alpha_m \varepsilon_{t-m}^2 + \sum_{n}^{Q} \beta_n \sigma_{t-n}^2 \quad (2)$$

In Equation 2, σ_t^2 is the conditional variance and $\varepsilon_t = \epsilon_t \sigma_t,\ \ \varepsilon_t \sim D(0, \sigma_t^2)$ defines the residual terms of ARMA*(p,q)* that are distributed according to the distribution *D*. Moreover, ω is the intercept, β_n and α_m are the parameters defining the conditional variance.

In this research, one-day ahead price direction predictions of ARMA*(p,q)* – GARCH*(P,Q)* is obtained as follows:

- Fit ARMA*(p,q)* – GARCH*(P,Q)* to window *i* $[w_{i,j} \in \{w_{1,j}, \dots, w_{50,j}\}]$. Where w_i represents a model fit window consisting 250 differenced closing price observations of Period *j* [$j \in \{1,2\}$].
- Determine the window specific ($w_{i,j}$) orders of $(p,q) \in \{0,1,2\}$, $P \in \{1,2,3\}$ and $Q \in \{0,1,2,3\}$ as well as the distribution $D \in$ {Normal, Student-t} according to the AIC criteria.
- Obtain one-day ahead forecasts for each window *i* and Period *j* ($w_{i,j}$).

Overall, ARMA*(p,q)* – GARCH*(P,Q)* is fitted 100 times to each index data by applying a one-day ahead rolling windows approach in order to obtain all the out-sample price direction predictions.

3.2. Ridge Regression

Proposed by Hoerl and Kennard (1970), Ridge regression is an extended form of linear regression that also incorporates ℓ_2 regularization into estimation. It minimizes the residual sum of squares of the regression with a restriction or penalty on the ℓ_2-norm of the coefficients. Instead of completely omitting *non-significant* variables, Ridge regression assigns coefficient values close to zero. Additionally, it penalizes highly correlated variables by shrinking their coefficients towards each other. Suppose Y is the vector of respondent (outcome or label) variable with *n* number of observations and X is the predictor variable matrix with *n* observations for each predictor *x* for a total number of *p* predictors, then the minimization problem estimating the coefficients of the Ridge regression can be written as (James *et al.*, 2013):

$$\hat{\beta}^{Ridge} = arg \min_{\beta} \left[\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2\right] \quad (3)$$

where $\lambda \geq 0$ is the regression hyper-parameter controlling the degree of parameter shrinkage. The bigger the lambda, the more the parameters are penalized. If $\lambda = 0$, then the solution is the OLS estimate of linear regression.

3.3. Support Vector Machines

As one of the supervised machine learning algorithms, Support Vector Machines (SVM) are rooted from the statistical learning theory or *VC theory* of Vapnik and Chervonenkis (1974) and Vapnik (1982, 1995) (Smola & Schölkopf, 2004). SVM algorithm can be used both for classification and regression tasks. Mainly, SVM is an algorithm which can linearly (in the simplest case) or non-linearly map an input vector into a high-dimensional feature space where an *optimal separating hyperplane* is defined (Vapnik, 1998). The points on the feature space, that are used to define the optimal hyperplane by determining the maximal margins (maximum distance) between the vectors of different classes, are called *support vectors* (Cortes & Vapnik, 1995).

In the simplest form, where the input data is linearly separable or classifiable in the original input space, the linear separating functions (support vectors) are the hyperplanes. On the other hand, if the data is not linearly separable, SVM uses a kernel function to non-linearly map input vectors (data) into the feature space (Kecman, 2005). As explained by Vapnik (1998), the complexity of constructing a SVM is associated with the number of support vectors rather than the dimension of the feature space. The following paragraphs focus only on the definition of SVM for the regression case, since in this research SVM is applied to forecast price directions of the indices including their magnitude additional to their sign.

More formally, let us assume that *x* is the *n*-dimensional input vector $\left(x_p \in R^n\right)$ and *y* is the output/label vector $\left(y_p \in R\right)$. SVM can be defined as:

$$\hat{y} = f(x, w) = w^T\phi(x) + b \quad (4)$$

where $\hat{y}$ is the output of SVM, $w \in F$ is the weight vector and the term *b* is defined as the bias. Moreover, $\phi(x)$ is the basis function that can non-linearly map the input vector into a high-dimensional feature space *F*. Compared to SVM for classification tasks in which margins is used, SVM for the regression case uses Vapnik's *ε-insensitivity loss function* that is defined as (Kecman, 2005):

$$\mathcal{L}_{\varepsilon} = |y - \hat{y}|_{\varepsilon} = \begin{cases} 0\,, & if \quad |y - \hat{y}| \leq \varepsilon \\ |y - \hat{y}| - \varepsilon\,, & otherwise \end{cases} \quad (5)$$

In order to estimate w and b coefficients of SVM, the following constrained function (ψ) is minimized:

$$\min\ \psi_{w,\xi,\xi^*} = \frac{1}{2}w^T w + C\sum_{p=1}^{n}\left(\xi_p + \xi_p^*\right) \quad (6)$$

Satisfying the below constraints for $p \in \{1,2, \dots, n\}$:

$$y_p - \hat{y}_p \leq \varepsilon + \xi_p \quad (7)$$

$$\hat{y}_p - y_p \leq \varepsilon + \xi_p^* \qquad \wedge \qquad \xi_p, \xi_p^* \geq 0 \quad (8)$$

where *C* is the penalization or cost parameter, ξ_p and ξ_p^* are the newly introduced *positive slack variables*. With the help of the Lagrange multipliers and optimality conditions, SVM defined in Equation 4 takes the form:

$$\hat{y}_p = f(x, w) = \sum_{p=1}^{n} \left(\alpha_p - \alpha_p^*\right) K\left(x_p, x\right) + b \tag{9}$$

In Equation 9, the difference of $\left(\alpha_p - \alpha_p^*\right)$ define the support vectors and $K\left(x_p, x\right)$ is the kernel function. SVM can use various kernel functions. In this research, linear as well as the Gaussian Radial Basis kernel functions are employed to obtain predictions of SVM.

3.4. Relevance Vector Machines

Proposed by Tipping (2001), Relevance Vector Machine (RVM) is also a supervised machine learning algorithm with a similar linear functional form (see Equation 4) as SVM. Apart from SVM, RVM incorporates probabilistic Bayesian learning into estimation to improve prediction accuracy of the outcome variable by considering model or prediction uncertainty. As stated by Tipping (2001), SVM is not always sparse and parsimonious enough resulting from several drawbacks, such as; an increasing number of basic functions and complexity due to an increase in training data; the Mercer's condition must be hold by the employed kernel functions. Moreover, in regression tasks SVM yields point estimates of the outcomes ignoring uncertainty of the prediction.

On the other hand, RVM does not carry any of the mentioned drawbacks of SVM. More formally, given the input and output vectors of $\left\{x_p, y_p\right\}_{p=1}^{N}$, RVM aims to predict the output *y* by (Tipping, 2001):

$$\hat{y} = f(x, w) + \varepsilon_p \tag{10}$$

where $\boldsymbol{w} = \left(w_0, \ldots, w_p\right)^T$ is the weight vector and $\varepsilon_p \sim N(0, \sigma^2)$ is a normally distributed error term with zero mean and variance σ^2. The likelihood of the observed training data is defined as:

$$p(y|w, \sigma^2) = (2\pi\sigma^2)^{-N/2}\ exp\left(-\frac{\|y - \phi w\|^2}{2\sigma^2}\right) \tag{11}$$

In Equation 11, the term ϕ represents an $Nx(N+1)$ matrix of $\phi = [\phi(x_1), \ldots, \phi(x_N)]^T$ and $\phi\left(x_p\right) = \left[1, K\left(x_p, x_1\right), \ldots, K\left(x_p, x_N\right)\right]^T$.

Under the Bayesian framework, RVM assigns a zero-mean Gaussian prior distribution over the weight parameters as follows:

$$p(w|\alpha) = \prod_{p=0}^{N} N\left(w_p|0, \alpha_p^{-1}\right) \tag{12}$$

On the other hand, the posterior distributions over the weights are defined as (Tipping, 2001):

$$p(w|y, \alpha, \sigma^2) = (2\pi)^{[(-N-1)/2]}\ |\Sigma|^{(-1/2)}\ \exp[(-1/2)\ (w-\mu)^T\ \Sigma^{-1}\ (w-\mu)] \tag{13}$$

where $\mu = (\sigma^{-2}\, \Sigma\, \phi^T y)$ is the posterior mean and $\Sigma = [\sigma^{-2}\phi^T\phi + A]^{-1}$ is the posterior covariance with $A = diag(\alpha_1, \ldots, \alpha_N)$.

Furthermore, in order to obtain uniform hyperpriors, the marginal likelihood $\left(p(y|\alpha, \sigma^2)\right)$ is maximized as in Equation 14.

$$p(y|\alpha, \sigma^2) = \int p(y|w, \sigma^2)\, p(w|\alpha) dw \tag{14}$$

Like SVM, in this research linear as well as the Gaussian Radial Basis kernel functions are employed to obtain predictions of RVM.

3.5. Performance Evaluation

Index direction prediction performance of the applied models is compared with the following performance metrics:

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt[2]{\frac{\sum_{k=1}^{T}(y_k - \hat{y}_k)^2}{T}} \tag{15}$$

where y_k is the out of sample observed values and $\hat{y}_k$ is the corresponding predictions of a model. T is the total number of out-of-sample observations.

Mean Absolute Error (MAE):

$$MAE = \frac{\sum_{k=1}^{T}|y_k - \hat{y}_k|}{T} \tag{16}$$

Moreover, to be able to observe the proportion of the variation of the respondent variable explained by each method, coefficient of determination (R^2) statistic is also estimated by:

$$R^2 = 1 - \frac{\sum_{k=1}^{T}(y_k - \hat{y}_k)^2}{\sum_{k=1}^{T}(y_k - \bar{y})^2} \tag{17}$$

where y_k, $\bar{y}$ and $\hat{y}_k$ are the out-of-sample observed *y*, mean of the observed *y* and predicted values of the respondent variable, respectively.

Additional to the above metrics that compare the real observations with the predicted values in terms of their magnitude, the following performance metrics, which are obtained by comparing the sign of the out-of-sample observations with the model predictions, are also estimated.

Accuracy shows the proportion of correctly predicted index direction ($\hat{y}$) in terms of its sign:

$$\text{Accuracy} = \frac{P^{true} + N^{true}}{T} \tag{18}$$

where P^{true} is the total number of positive direction predictions conditional on having positive out-of-sample realizations. Similarly, N^{true} is the total number of negative direction predictions when the realized observations are also negative. The confusion matrix shown in Table 1 explains the terms clearly.

Table 1: Confusion Matrix

		Predicted	
		Positive	**Negative**
Actual	**Positive**	P^{true}	N^{false}
	Negative	P^{false}	N^{true}

Precision defines the proportion of correctly predicted positive directions over all the positive direction predictions.

$$\text{Precision} = \frac{P^{true}}{P^{true} + P^{false}} \tag{19}$$

On the other hand, Recall is the proportion of correctly predicted positive directions over all the positive direction realizations (positive out-of-sample observations).

$$\text{Recall} = \frac{P^{true}}{P^{true} + N^{false}} \quad (20)$$

F-1 score is estimated from the obtained Precision and Recall values as follows:

$$\text{F-1 Score} = \frac{2 \ x \ Precision \ x \ Recall}{Precision + Recall} \quad (21)$$

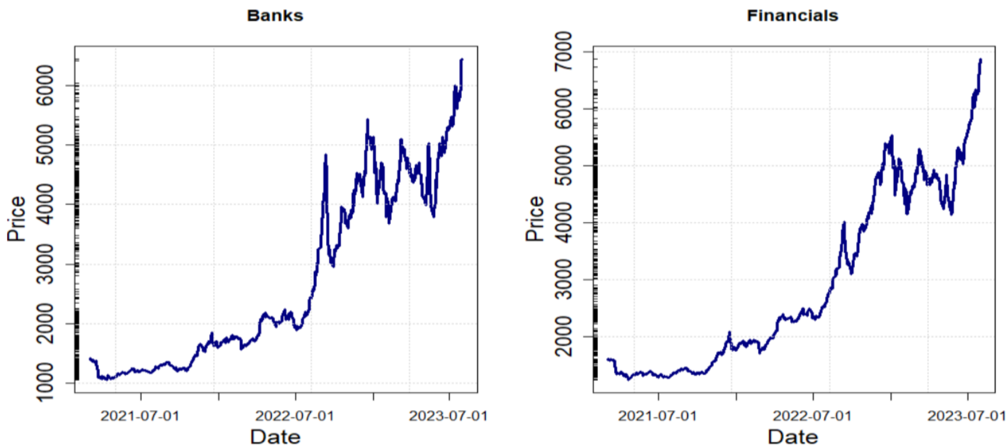
Mainly, F-1 score is employed to compare two forecasting models when one of them indicates a low Precision and the other shows a high Recall. A model has its maximum F-1 score when its Precision is equal to its Recall (Sahu *et al.*, 2023).

4. Data and Empirical Findings

4.1. Data

This research employs daily closing prices of two stock indices of Borsa Istanbul (BIST Banks and the BIST Financials) obtained from the Investing website for the period of 03 March 2021 - 31 July 2023. The first index, BIST Banks (XBANK) is a market value weighted price index consisting twelve banks listed in Istanbul Stock Exchange. On the other hand, BIST Financials (XUMAL) consists 129 companies (as of 24 September 2023) operating in various industries, such as banks, holdings, and real estate investment trusts (Borsa Istanbul, 2023). Time development of the two indices is given in Figure 1.

Figure 1: Daily Closing Prices of BIST Banks and BIST Financials



Source: Author's own illustration

Descriptive statistics of the price series are reported in Table 2. According to the obtained descriptives, both series are neither stationary nor normally distributed. Moreover, Ljung-Box (LB) (Ljung & Box, 1978) and Lagrange Multiplier (LM) (Engle, 1984) tests applied for lags 7 and 12, respectively indicate existence of auto-correlation and heteroskedasticity.

Table 2: Descriptive Statistics of the Indices – Full Sample

Index	**Mean**	**Stand. Dev.**	**Jarque-Bera (p-val.)**	**ADF (p-val.)**	**LB (Q12) (p-val.)**	**LM (Q5) (p-val.)**	**LM (Q12) (p-val.)**
Banks	2686	1469.6	67.20 (0.00)	-2.17 (0.50)	6545.8 (0.00)	589.15 (0.00)	582.46 (0.00)
Financials	2891	1513.8	64.99 (0.00)	-1.30 (0.87)	6658.8 (0.00)	592.2 (0.00)	585.36 (0.00)

4.2. Data Pre-Processing

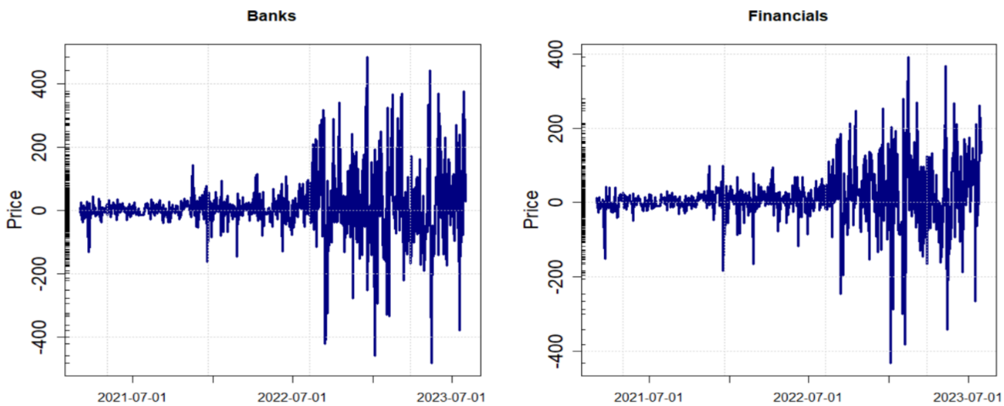
In order to predict price direction of the indices, daily closing prices are differenced from their previous observations as follows:

$$y_t' = y_t - y_{t-1} \tag{22}$$

where y_t is the closing price of an index observed at time t and y_t' is new price series obtained following the differencing transformation. Time development of the transformed series is given in Figure 2.

Following the differencing transformation, price series are divided into two main periods in which each ML algorithm is trained and tested separately. The first period starts from 03 March 2021 till 16th of May 2022 and the second period starts from 17 May 2022 till 31th of July 2023. Period specific descriptive statistics of the differenced series (y_t') are given in Table 3.

Figure 2: Differenced Closing Prices of BIST Banks and BIST Financials



Source: Author's own illustration

According to Table 3, the mean and standard deviation of both series are considerably smaller in Period 1 compared to Period 2 which can also be observed from Figure 2. In both periods the transformed series are not normally distributed but are stationary. For both indices, while heteroskedasticity is prevalent in Period 1 and 2, auto-correlation is not observed in Period 1.

Table 3: Descriptive Statistics of the Differenced Series – Period 1 and Period 2

Index	Mean	Standard Dev.	Jarque-Bera (p-value)	ADF (p-value)	LB (Q12) (p-value)	LM (Q5) (p-value)
			Period 1			
Banks	1.89	32.05	466.6 (0.00)	-5.42 (0.01)	20.38 (0.06)	18.34 (0.003)
Financials	2.19	30.47	1303 (0.00)	-5.71 (0.01)	13.68 (0.32)	13.06 (0.023)
			Period 2			
Banks	14.91	146.30	29.82 (0.00)	-6.82 (0.01)	25.03 (0.015)	32.81 (0.000)
Financials	15.39	106.65	74.90 (0.00)	-6.51 (0.01)	24.56 (0.017)	37.87 (0.000)

Source: Author

Furthermore, 83% of the differenced data of each period is used for algorithm training and the rest of the 17% (unseen data) is employed for testing the prediction performance of the employed ML methods. Before training the algorithms, the commonly employed min-max normalization is applied to the series in order to be able enhance ML algorithm's learning. Moreover, for a purpose to prevent data leakage, period specific train and test data is normalized with the minimum and maximum values obtained from the train data only.

$$y_{t,w}^{n} = \left(y_{t,w} - \min\left(y_{t,w}^{tr}\right)\right) / \left(\max\left(y_{t,w}^{tr}\right) - \min\left(y_{t,w}^{tr}\right)\right) \tag{23}$$

where $y_{t,w}^{n}$ is the period specific $w \in \{1,2\}$ normalized price series and $y_{t,w}^{tr}$ is the period specific train data. One-day ahead closing price directions of each index is predicted by employing five-day lagged values of their own observations as follows:

$$\hat{y}_t' = f(y_{t-1}', y_{t-2}', y_{t-3}', y_{t-4}', y_{t-5}') + \varepsilon_t \tag{24}$$

4.3. Algorithm Calibration

Hyper-parameters of the ML algorithms are trained and calibrated by using the time series cross-validation approach (Hyndman & Athanasopoulos, 2018). Compared to commonly employed k-fold cross-validation, time series cross-validation (ts-cv) preserves the order of the data which enables the algorithms to model time series specific patterns that are highly prevalent in financial series. Parameters of all the three ML algorithms are tuned with the same ts-cv approach which employed a one-day ahead rolling forecasting resampling method (window size is 100 daily observations), to validate a pre-specified forecasting horizon that is 1 day in this case (h=1).

Following the training, model specific hyper-parameters that minimize the model RMSE values are selected. For this purpose, *train* function of the *caret* package (Kuhn, 2008) of R software (R Core Team, 2019) is employed. Period and model specific selected optimal parameters of ML algorithms are summarised in Table 4.

Table 4: Hyper-Parameters of Ridge Regression, SVM and RVM

		Banks		Financials	
Model	**Parameter**	**Period 1**	**Period 2**	**Period 1**	**Period 2**
Ridge Regression	Lambda	0.180	10	3.554	9.099
SVM	Cost	0.010	0.010	0.010	0.010
SVM_{radial}	Sigma / Cost	0.06 / 0.12	0.72 / 0.06	0.1 / 0.09	0.01 / 0.1
RVM_{radial}	Sigma	0.080	0.050	0.010	0.040

Note: While SVM uses the linear kernel, SVM_{radial} and RVM_{radial} use the Gaussian Radial Basis kernel functions.

Following the hyper-parameter selection, one-day ahead predictions of the algorithms are obtained for each out-of-sample test period.

4.4. Results

Once, one-day ahead direction predictions of the models are obtained, their performance is compared with various metrics explained in 3.5. Performance Evaluation section. Table 5 and Table 6 summarize the estimated performance metrics for BIST Banks and BIST Financials, respectively.

Table 5: Performance Metrics of BIST Banks

Period	Model	RMSE**	R^2	MAE**	Accuracy	Precision	Recall	F1-score
1	ARMA-GARCH	0.1241	0.0539	0.0901	56%	57.14%	74.07%	64.52%
	Ridge Regression	0.1533	**0.0658**	0.1247	48%	*60.00%*	11.11%	18.75%
	SVM*	0.1196	0.0000	**0.0859**	48%	51.72%	55.56%	53.57%
	SVM_{radial}	*0.1185*	0.0030	0.0866	56%	57.14%	74.07%	64.52%
	RVM*	**0.1175**	*0.0618*	0.0875	**62%**	**62.50%**	74.07%	**67.80%**
	RVM_{radial}	0.1195	0.0018	*0.0863*	*58%*	57.89%	**81.48%**	*67.69%*
2	ARMA-GARCH	**0.1547**	**0.0302**	**0.1207**	*58%*	*63.64%*	*70.00%*	*66.67%*
	Ridge Regression	0.5343	0.0005	0.5120	40%	0.00%	0.00%	0.00%
	SVM*	0.1594	0.0008	0.1267	52%	59.38%	63.33%	61.29%
	SVM_{radial}	0.1601	0.0000	0.1271	40%	50.00%	30.00%	37.50%
	RVM*	0.1770	0.0028	0.1441	50%	58.06%	60.00%	59.02%
	RVM_{radial}	*0.1557*	*0.0209*	*0.1228*	**62%**	**64.86%**	**80.00%**	**71.64%**

Notes: *The algorithm uses the linear kernel function. **Estimated from the normalized out-of-sample predictions. The best and the second-best performers in each metric are shown in bold and italics, respectively.

As can be seen from Table 5, the first three metrics of RMSE, R^2 and MAE rank different models as the best depending on the metric in the first test period of BIST Banks. While Ridge regression and SVM with the linear kernel had the best R^2 and MAE metrics, respectively, RVM with the linear kernel yielded the best RMSE and the second best R^2 of Period 1. Additionally, RVM obtained the highest Accuracy, Precision, and F1-scores indicating the algorithm's ability to correctly predict the highest proportion of all direction realizations (positive and negative) with a good precision in positive direction predictions. Moreover, RVM_{radial} obtained the highest recall percentage indicating the highest number of correctly predicted positive directions over all the positive direction realizations. Compared to Period 1, the metrics estimated in Period 2 of BIST Banks rank mainly two models as the best. While RMSE, R^2 and MAE rank the traditional ARMA-GARCH as the best and RVM_{radial} as the second best, all the rest four metrics indicate an outperformance of RVM_{radial}.

Overall, even though there are some period specific discrepancies in the first three metrics, the last four metrics (Accuracy, Precision, Recall and F1-score) that are estimated by classifying the out-of-sample model predictions depending on their sign (positive or negative) are unanimous in their decision by ranking RVM as the best in both periods of BIST Banks. While all the metrics are estimated from out-of-sample model predictions, the last four are reported in percentages (metric x 100). Moreover, among the applied prediction models, Ridge Regression yielded the worst RMSE and MAE values as well as F1-score in both periods of BIST Banks.

Table 6: Performance Metrics of BIST Financials

Period	Model	RMSE**	R^2	MAE**	Accuracy	Precision	Recall	F1-score
1	ARMA-GARCH	0.1135	0.0004	0.0848	*64%*	*66.67%*	90.91%	*76.92%*
	Ridge Regression	0.5003	**0.0172**	0.4881	34%	0.00%	0.00%	0.00%
	SVM*	*0.1125*	0.0010	*0.0837*	*64%*	*66.67%*	90.91%	*76.92%*
	SVM_{radial}	**0.1117**	0.0020	**0.0825**	**70%**	**68.75%**	**100.00%**	**81.48%**
	RVM*	0.1171	*0.0133*	0.0933	56%	64.86%	72.73%	68.57%
	RVM_{radial}	0.1149	0.0019	0.0860	62%	64.58%	*93.94%*	76.54%
2	ARMA-GARCH	0.1444	0.0008	*0.1165*	62%	65.85%	84.38%	73.97%
	Ridge Regression	0.5556	0.0141	0.5388	36%	0.00%	0.00%	0.00%
	SVM*	**0.1434**	0.0229	**0.1147**	**68%**	69.05%	*90.63%*	**78.38%**
	SVM_{radial}	0.1447	**0.0289**	0.1166	*66%*	66.67%	**93.75%**	*77.92%*
	RVM*	0.1490	*0.0250*	0.1216	64%	*70.59%*	75.00%	72.73%
	RVM_{radial}	*0.1438*	0.0183	0.1171	**68%**	**73.53%**	78.13%	75.76%

Notes: *The algorithm uses the linear kernel function. **Estimated from the normalized out-of-sample predictions. The best and the second-best performers in each metric are shown in bold and italics, respectively.

When the BIST Financials predictions of the models are evaluated, SVM_{radial} is ranked as the best in most of the (six out of seven) estimated performance metrics in Period 1. The exception is the R^2 metric in which Ridge regression and RVM with the linear kernel obtained the best and the second-best values, respectively.

In Period 2, SVM that uses the linear kernel is outperformed the rest in four out of seven metrics. Moreover, RVM_{radial} yielded the highest out-of-sample Accuracy and Precision percentages in Period 2. Even though, the Accuracy percentages of SVM and RVM_{radial} are equal (68%) indicating an equal proportion of correctly predicted directions (the total of positive and negative) over all the out-sample realizations, the F1-score points a better positive direction prediction performance in favour of SVM. Similar to the metrics of BIST Banks, Ridge Regression yielded the worst RMSE, MAE and F1-score in both periods of BIST Financials too. Furthermore, the traditional ARMA-GARCH is not ranked as the best in any of the metrics of BIST Financials in both periods.

5. Conclusion

This research assessed and compared index direction prediction performance of RVM with the ARMA-GARCH, Ridge Regression and the commonly employed ML algorithm of SVM for two underexplored indices of Borsa Istanbul. The out-of-sample direction predictions of the models are compared by using seven different and commonly employed performance metrics.

According to the obtained results, the first three metrics of RMSE, R^2 and MAE are not able differentiate and rank a specific prediction model as the best unanimously, except the second period of BIST Banks index. Since these metrics are estimated from the point forecasts and especially one of the metrics, RMSE, is a scale-dependent accuracy measure that tends to be highly affected from the outliers (Hyndman & Koehler, 2006), these results are not surprizing. The surprizing results came from the Ridge Regression where it ended up with the worst RMSE, MAE and F1-score in both periods of both indices indicating a consistent bad performance. As mentioned by James *et al.* (2013) performance of the linear penalized regressions, like Ridge regression, is dependent on the data and the feature relation types that one aims to predict. Difficulty arises from trying to predict nonlinear associations with a linear model. Moreover, it is also difficult to find research papers that focus on the indices of

Borsa Istanbul and compare the price and/or direction prediction performance between the linear Ridge regression and nonlinear ML algorithms which can be used as a comparison for this research paper.

On the other hand, RVM yielded most of the best metrics in both periods of the BIST Banks index as well as the highest Accuracy and Precision percentages in the second period of BIST Financials. Additionally, compared to RVM with the linear kernel, RVM that use the Gaussian Radial Basis kernel performed better in the second period of both indices in which the mean and volatility of the series were higher. While there are numerous works assessing price and/or direction prediction of SVM using either BIST100 or BIST30 indices of Borsa Istanbul (see for example; Kara *et al.* (2011), Oztekin *et al.* (2016), Pabuçcu (2019) and Kartal (2020)), there is none employing the BIST Banks and Financials indices as well as the RVM algorithm. Considering the SVM algorithm, the classification accuracies of BIST indices reported in the literature relies between 50% to 90% depending on the analysis period and the algorithm. The findings of this paper are in line with the findings of literature employing the SVM and at least one of the indices of Borsa Istanbul.

Overall, the results of this research show the applicability of RVM in predicting direction of two of the financial indices of Borsa Istanbul with a potential of the algorithm to beat a high performer like SVM depending on the period and the kernel function.

References

Akcan, A., & Kartal, C. (2011), “İMKB Sigorta Endeksini Olusturan Sirketlerin Hisse Senedi Fiyatlarının Yapay Sinir Ağları İle Tahmini”, Muhasebe Ve Finansman Dergisi, 51: 27-40.

Aksoy, B. (2021), “Pay senedi fiyat yönünün makine öğrenmesi yöntemleri ile tahmini: Borsa İstanbul örneği”, Business and Economics Research Journal, 12(1): 89-110. http://dx.doi.org/10.20409/berj.2021.312

Avcı, E. (2007), “FORECASTING DAILY AND SESSIONAL RETURNS OF THE ISE-100 INDEX WITH NEURAL NETWORK MODELS”, Doğuş Üniversitesi Dergisi, 8(2): 128-142.

Ballings, M., Poel, D.V., Hespeels, N., & Gryp, R. (2015), “Evaluating multiple classifiers for stock price direction prediction”, Expert Systems with Applications, 42(20): 7046–7056. https://doi.org/10.1016/j.eswa.2015.05.013

Bollerslev, T. (1986), “Generalized Autoregressive Conditional Heteroskedasticity”, J. Econ., 31: 307-327.

Borsa Istanbul (2023), https://www.borsaistanbul.com/ (Access date: 25.09.2023).

Boyacıoğlu, M.A., & Avcı, D. (2010), “An Adaptive Network-Based Fuzzy Inference System (ANFIS) for the prediction of stock market return: The case of the Istanbul Stock Exchange”, Expert Systems with Applications, 37: 7908–7912. https://doi.org/10.1016/j.eswa.2010.04.045

Cortes, C., & Vapnik, V. (1995), “Support-vector networks”, Machine Learning, 20(3): 273-297.

Diler, A.I. (2003), “Predicting direction of ISE national-100 index with backpropagation trained neural network”, Journal of Istanbul Stock Exchange, 7(25–26): 65–81.

Engle, R. (1982), “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation”, Econometrica, 50(4): 987-1007.

Engle, R. (1984), “Chapter 13 Wald, likelihood ratio, and Lagrange multiplier tests in econometrics”, Handbook of Econometrics, North Holland, 2: 775–826.

Filiz, E., Akogul, S., & Karaboğa, H.A. (2021), “Büyük Dünya Endeksleri Kullanılarak BIST-100 Endeksi Değişim Yönünün Makine Öğrenmesi Algoritmaları ile Sınıflandırılması”, Bitlis Eren Üniversitesi Fen Bilimleri Dergisi, 10(2): 432–441. https://doi.org/10.17798/bitlisfen.889007

Gündüz, H., Çataltepe, Z., & Yaslan, Y. (2017) "Stock daily return prediction using expanded features and feature selection", Turkish Journal of Electrical Engineering and Computer Sciences, 25(6): 4829-4840. https://doi.org/10.3906/elk-1704-256

Henrique, B.M., Sobreiro, V.A., & Kimura, H. (2019), “Literature review: Machine learning techniques applied to financial market prediction”, Expert Systems with Applications, 124: 226-251. https://doi.org/10.1016/j.eswa.2019.01.012

Hoerl, A., & Kennard, R. (1970), “Ridge Regression: Biased Estimation for Nonorthogonal Problems”, Technometrics, 12(1): 55-67. https://doi.org/10.2307/1267351

Huang, S.-C., & Wu, T.-K. (2008), “Combining wavelet-based feature extractions with relevance vector machines for stock index forecasting”, Expert Systems, 25: 133-149. https://doi.org/10.1111/j.1468-0394.2008.00443.x

Hyndman, R.J., & Athanasopoulos, G. (2018), Forecasting: principles and practice (2. ed.). Retrieved from: OTexts.com/fpp2. (Access date: 01.03.2021).

Hyndman, R.J., & Koehler, A.B. (2006), “Another look at measures of forecast accuracy”, International Journal of Forecasting, 22(4): 679-688. https://doi.org/10.1016/j.ijforecast.2006.03.001

https://tr.investing.com (Access date: August 25, 2023).

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013), An Introduction to Statistical Learning with Applications in R: Springer.

Kara, Y., Boyacioglu, M.A., & Baykan, Ö.K. (2011), “Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange”, Expert Systems with Applications, (38): 5311-5319.

Kartal, C. (2020). “Destek Vektör Makineleri ile Borsa Endekslerinin Tahmini”, İnsan ve Toplum Bilimleri Araştırmaları Dergisi, 9(2): 1394-1418.

Kecman, V. (2005), “Support vector machines–an introduction”, Support vector machines: Theory and applications (pp. 1-47), Springer.

Kılıç, A., Güloğlu, B., Yalçın, A., & Üstündağ, A. (2023), “Big data–enabled sign prediction for Borsa Istanbul intraday equity prices”, Borsa Istanbul Review, 23(2): 38-52. https://doi.org/10.1016/j.bir.2023.08.005

Kuhn, M. (2008), “Building Predictive Models in R Using the caret Package”, Journal of Statistical Software, 28(5): 1–26. https://doi.org/10.18637/jss.v028.i05

Kumbure, M.M., Lohrmann, C., Luukka, P., & Porras, J. (2022), “Machine learning techniques and data for stock market forecasting: A literature review”, Expert Systems with Applications, 197: 116659. https://doi.org/10.1016/j.eswa.2022.116659

Ljung, G.M., & Box, G.E.P. (1978), “On a measure of lack of fit in time series models”, Biometrika, 65: 297–303.

Nikou, M., Mansourfar, G., & Bagherzadeh, J. (2019), “Stock price prediction using DEEP learning algorithm and its comparison with machine learning algorithms”, Intell Sys Acc Fin Mgmt, 26: 164–174. https://doi.org/10.1002/isaf.1459

Özgür, C., & Sarıkovanlık, V. (2022), “FORECASTING BIST100 AND NASDAQ INDICES WITH SINGLE AND HYBRID MACHINE LEARNING ALGORITHMS”, Economic Computation and Economic Cybernetics Studies and Research, 3(56): 235-250. https://doi.org/10. 24818/18423264/56.3.22.15

Oztekin, A., Kizilaslan, R., Freund, S., & Iseri, A. (2016), “A Data Analytic Approach to Forecasting Daily Stock Returns in an Emerging Market”, European Journal of Operational Research, 253(3): 697-710. https://doi.org/10.1016/j.ejor.2016.02.056

Pabuçcu, H. (2019), “Borsa endeksi hareketlerinin makine öğrenme algoritmaları ile tahmini”, Uluslararası İktisadi ve İdari İncelemeler Dergisi, (23): 179-190. https://doi.org/10.18092/ulikidince.484138

R Core Team (2019), “R: A Language and Environment for Statistical Computing: R Foundation for Statistical Computing”. Retrieved from: https://www.r-project.org/

Sahu, S.K., Mokhade, A., & Bokde, N.D. (2023), “An Overview of Machine Learning, Deep Learning, and Reinforcement Learning-Based Techniques in Quantitative Finance: Recent Progress and Challenges”, Applied Sciences, 13: 1956. https://doi.org/10.3390/app13031956

Smola, A.J., & Schölkopf, B. (2004), “A tutorial on support vector regression”, Statistics and Computing, 14: 199–222.

Tipping, M.E. (2001), “Sparse Bayesian learning and the relevance vector machine”, Journal of Machine Learning Research, 1: 211–244.

Ünlü, K.D., Potas, N., & Yılmaz, M. (2021), “Forecasting Direction of BIST 100 Index: An Integrated Machine Learning Approach”, Chaos, Complexity and Leadership 2020, Springer Proceeding in Complexity, Springer. https://doi.org/10.1007/978-3-030-74057-3_5

Vapnik, V. (1995), The Nature of Statistical Learning Theory, Berlin: Springer Science & Business Media.

Vapnik, V. (1998), Statistical Learning Theory, Toronto: John Wiley & Sons.

Vapnik, V., & Chervonenkis A. (1974), Theory of Pattern Recognition [in Russian]. Nauka, Moscow. (German Translation: Wapnik W. & Tscherwonenkis A., Theorie der Zeichenerkennung, Akademie-Verlag, Berlin, 1979).

Vapnik V.N. (1982), Estimation of Dependences Based on Empirical Data, Berlin: Springer.

Yümlü, S., Gürgen, F.S., & Okay, N. (2005), “A comparison of global, recurrent and smoothed-piecewise neural models for Istanbul stock exchange (ISE) prediction”, Pattern Recognition Letters, 26(13): 2093-2103. https://doi.org/10.1016/j.patrec.2005.03.026